COMS W6998-5: Algorithms through Geometric Lens (Fall'17)	Oct 9, 2017
Lecture 10 – Locality-Sensitive Hashing	
Instructors: *Alex Andoni, Ilya Razenshteyn*	Scribes: *Nicholas Herrera*

1 Introduction

Today's lecture finishes the discussion of locality-sensitivity hashing (LSH) and introduces two graph-theoretic problems – MAX-CUT and graph colorings. The main problem of this lecture is stating and analyzing an optimal construction of LSH for $(S^{d-1}, \|\cdot\|_2)$.

1.1 Review from Last Lecture

Definition 1 (LSH)**.** *A random partition $\mathcal{P}$ is a locality sensitive hashing on a space X with parameters (p_1, p_2) if*

- *for all $x, y \in X$ such that $\|x - y\| \le r$, we have that* $\Pr[\mathcal{P}(x) = \mathcal{P}(y)] \ge p_1$
- *for all $x, y \in X$ such that $\|x - y\| > cr$, we have that* $\Pr[\mathcal{P}(x) = \mathcal{P}(y)] \le p_2$.

Theorem 2. *If there exists an "efficient" LSH scheme on a space, then there also exists a data structure for (c, r)-ANN such that*

- *space requirement is $\mathcal{O}(n^{1+\rho}/p_1 + nd)$*
- *query time is $\mathcal{O}(dn^\rho/p_1)$*
- *Probability of success is 0.9*

where $\rho = \frac{\log(1/p_1)}{\log(1/p_2)}$.

We looked at two instantiations of LSH:

1. Hamming coordinate sampling, where we achieved $\rho \le \frac{1}{c}$, which is optimal
2. The random hyperplane construction for LSH on $(S^{d-1}, \|\cdot\|_2)$, where we achieved $\rho \le \frac{1}{c}$ which is not optimal.

1.2 Goal for Today

Our goal for today is develop an optimal LSH construction for $(S^{d-1}, \|\cdot\|_2)$ with $\rho \le \frac{1}{c^2}$. This difference in ρ by a quadratic factor gives difference in query time for, say $c = 2$, of $\sqrt{n}$ time and $n^{1/4}$ time.

2 Candidate Constructions

We consider three similar candidate constructions for this LSH problem. Of these, we will analyze only LSH Family 3.

2.1 LSH Family 1

- Sample $u_1, u_2, \ldots, u_T \in S^{d-1}$ uniformly at random and independently
- Choose $h(p) = \text{argmin}_i \|u_i - p\|_2$

The partitions here are therefore given by the Voronoi diagram on the sphere given by the sampled vectors. We note that for $T = 2$ this construction is indentical to the random hyperplane construction.

Theorem 3. *In LSH Family 1, $\rho \leq \frac{1}{c^2} + \epsilon(T)$, where $\epsilon(T) \to 0$ as $T \to \infty$.*

2.2 LSH Family 2

- Sample $u_1, u_2, \ldots, u_T, \ldots \in S^{d-1}$ uniformly at random and independently
- Set parameter $\eta \geq 0$
- Choose $h(p) = \min_i$ such that $\langle u_i, p \rangle \geq \eta$.

Interpreted geometrically, each set of points p such that $\langle u_i, p \rangle \geq \eta$ is a ball (a spherical cap) on S^{d-1} of measure on the order of $e^{-\eta^2 d}$. So we need around $e^{\eta^2 d}$ balls to cover the ball. We will call this kind of procedure *ball carving.*

Theorem 4. *In LSH Family 2, $\rho \leq \frac{1}{c^2} + \epsilon(\eta)$, where $\epsilon(\eta) \to 0$ as $\eta \to 1$.*

2.3 LSH Family 3

- Sample $g_1, g_2, \ldots, u_T, \ldots \sim \mathcal{N}(0,1)^{\otimes d}$ independently
- Set parameter $\eta \geq 0$
- Choose $h(p) = \min_i$ such that $\langle g_i, p \rangle \geq \eta$.

The difference between LSH Family 2 and LSH Family 3 is that in 3 we sample the vectors from the d-dimensional standard Gaussian (i.e., over all of $\mathbb{R}^d$) rather than uniformly over the sphere.

3 Analysis of LSH Family 3

Goal: Compute Pr[h(x) = h(y)] for $x, y \in S^{d-1}$ such that $\|x - y\|_2 = s > 0$.

The event that $h(x) = h(y)$ is the event that the first gaussian vector g_i sampled such that either $\langle g_i, x \rangle \geq \eta$ or $\langle g_i, y \rangle \geq \eta$ is the same first such vector g_i for both, so

$$\Pr[h(x) = h(y)] = \frac{\Pr_{g \sim \mathcal{N}(0,1)^{\otimes d}}[\langle x, g \rangle \geq \eta \text{ and } \langle y, g \rangle \geq \eta]}{\Pr_{g \sim \mathcal{N}(0,1)^{\otimes d}}[\langle x, g \rangle \geq \eta \text{ or } \langle y, g \rangle \geq \eta]}.$$

3.1 Analysis of Denominator

To give an estimate on the denominator, we use the following claim without proof:

Claim 5. *For $X \sim \mathcal{N}(0,1)$,*

$$\Pr[X \geq t] \sim \frac{1}{t^2} \frac{e^{-t^2/2}}{\sqrt{2\pi}}.$$

Combining the union bound with this claim, we have that

$$\Pr[\langle x, g\rangle \geq \eta \text{ or } \langle y, g\rangle \geq \eta] = \Theta(1) \cdot \Pr[\langle x, g\rangle \geq \eta] = \Theta\left(\frac{1}{\eta}\right) e^{-\eta^2/2}.$$

3.2 Analysis of Numerator

By rotational invariance of the Gaussian measure, we can assume without loss generality that $x = e_1 = (1, 0, 0, \ldots, 0)$ and $y = \alpha e_1 + \beta e_2 = (\alpha, \beta, 0, \ldots, 0) \in S^{d-1}$.

Let φ be the angle between x and y. Then we have the following identities:

- $\alpha = \sin \varphi$
- $\beta = \cos \varphi$
- $\alpha = 1 - \frac{s^2}{2}$
- $\beta = \sqrt{1 - \alpha^2}$

We can then rewrite the numerator as

$$\Pr_{g \sim \mathcal{N}(0,1)^{\otimes d}}[\langle x, g\rangle \geq \eta \text{ and } \langle y, g\rangle \geq \eta] = \Pr_{x,y \sim \mathcal{N}(0,1)}[x \geq \eta \text{ and } \alpha x + \beta y \geq \eta].$$

From a geometric perspective, our goal is to estimate the size (under the Gaussian measure) of the region of points in the plane $(x, y) \in \mathbb{R}^2$ such that $x \geq \eta$ and $\alpha x + \beta y \geq \eta$. Let $\Delta = \Delta(\eta, \alpha, \beta)$ be the distance from $(0, 0)$ to this region.

Claim 6.

$$\underset{x,y \sim \mathcal{N}(0,1)}{Pr}[x \geq \eta \textit{ and } \alpha x + \beta y \geq \eta] \sim e^{-\eta^2/2}$$

To understand why this claim holds, we look at the lower and upper bounds separately.

For the upper bound holds, we prove a more general claim:

Claim 7. *Let A be a measurable subset of $\mathbb{R}^2$. Then,*

$$\underset{x,y \sim \mathcal{N}(0,1)}{Pr}[(x, y) \in A] \leq e^{-\Delta(0,A)^2/2},$$

where $\Delta(0, A)$ denotes the Euclidean distance from $(0, 0)$ to A.

Proof. We note that the measure of set A is at most that of $\mathbb{R}^2$ minus the ball center at $(0, 0)$ of radius $\Delta(0, A)$. Then, using a transformation into polar coordiantes, we compute

$$\Pr_{x,y \sim \mathcal{N}(0,1)}[(x, y) \in A] \leq \frac{1}{2\pi} \int_{\Delta(0,A)}^{\infty} \int_0^{2\pi} e^{-r^2/2}\ r\ dr\ d\varphi = e^{-\Delta(0,A)^2/2}$$

□

For lower bound, we claim that

Claim 8. $\Pr_{x,y\sim\mathcal{N}(0,1)}[x \geq \eta \text{ and } \alpha x + \beta y \geq \eta] \geq \frac{1}{poly(\Delta)} \cdot e^{-\Delta^2/2}$

Rather than prove this formally, we give an informal explanation. Namely, since the tails of the Gaussian distribution decay so rapidly, we should expect that only points in a small neighborhood of the intersection point of $X = \eta$ and $\alpha X + \beta y = \eta$ determine the probability (up to polynomial factors).

To put these bounds together, we note that

$$\Delta^2 = \eta^2 \left(1 + \frac{(1-\alpha)^2}{\beta^2}\right) = \eta^2 \frac{2(1-\alpha)}{\beta}$$

and so

$$\Pr_{x,y\sim\mathcal{N}(0,1)}[x \geq \eta \text{ and } \alpha x + \beta y \geq \eta] \in \left[\frac{1}{\text{poly}(\eta)}, 1\right] \cdot e^{-\eta^2 \frac{1-\alpha}{\beta}}.$$

To see that this probability for the numerator is reasonable, we perform a few sanity checks:

- If x and y are very close ($s \to 0$, $\alpha = 1 - \frac{\epsilon}{2}$, $\beta = \epsilon$), then we have a probability of around $e^{-\eta^2/2}$.
- If x and y are orthogonal ($s = \sqrt{2}$, $\alpha = 0$, $\beta = 1$), then we have a probability of around $e^{-\eta^2}$.
- If x and y are at antipodal points of the sphere ($s = 2$, $\alpha = 1 - \frac{\epsilon}{2}$, $\beta = \epsilon$), then we have a probability near 0.

3.3 Putting the analysis together

Putting together the numerator and denominator, we have that

$$\Pr[h(x) = h(y)] = \frac{\Pr_{g\sim\mathcal{N}(0,1)^{\otimes d}}[\langle x, g\rangle \geq \eta \text{ and } \langle y, g\rangle \geq \eta]}{\Pr_{g\sim\mathcal{N}(0,1)^{\otimes d}}[\langle x, g\rangle \geq \eta \text{ or } \langle y, g\rangle \geq \eta]} \in \frac{\left[\frac{1}{\text{poly}(\eta)}, 1\right] \cdot e^{-\eta^2 \frac{1-\alpha}{\beta}}}{\Theta\left(\frac{1}{\eta}\right) e^{-\eta^2/2}} \subseteq \left[\frac{1}{\text{poly}(\eta)}, \text{poly}(\eta)\right] \cdot e^{-\eta^2 \frac{1-\alpha}{2\beta}}.$$

So

$$\frac{1}{\log \Pr[h(x) = h(y)]} = \frac{\eta^2}{2} \frac{(1-\alpha)^2}{\beta} + \varepsilon(\eta) \approx \frac{\eta^2}{2} \frac{(1-\alpha)^2}{\beta} = \frac{\eta^2}{2} \frac{s^2}{4 - s^2}$$

Then,

$$\begin{aligned} \rho &= \frac{\log(1/p_1)}{\log(1/p_2)} \\ &= \frac{\frac{\eta^2}{2}\frac{r^2}{4-r^2}}{\frac{\eta^2}{2}\frac{c^2r^2}{4-c^2r^2}} + \dots \\ &= \frac{1}{c^2} \frac{4 - c^2r^2}{4 - r^2} + \dots \\ &\leq \frac{1}{c^2} + \dots, \end{aligned}$$

where the ellipses are error terms that tend to 0 as η grows.

4 MAX-CUT and Graph Coloring

We consider two graph problems that at first blush appear to have nothing to do with geometry:

Problem 1 (MAX-CUT)**.** *Given an undirected graph $G = (V, E)$ with n vertices and m edges, find a partition $V = V_1 \sqcup V_2$ such that $E(V_1, V_2)$, the number of edges from V_1 to V_2, is maximized.*

For example, in a biparite graph, the MAX-CUT is the partition of the vertices into the two independent sets, giving a cut through all m edges. However, MAX-CUT is NP-hard in general, so we want to find a good approximation algorithm for it.

Problem 2. *(Coloring) Let $G = (V, E)$ be an undirected 3-colorable graph. Find a 3-coloring for G.*

This problem is NP-hard as well. How many colors do we need to use to be able to color a 3-colorable graph efficiently?

Our approach to designing approximation algorithms problems is as follows. We find an embedding of the graph G into S^{d-1} using a semidefinite programming relaxtion. We then use LSH to get a partition.

Theorem 9. *There exists a polynomial time algorithm that finds a cut $V = V_1 \sqcup V_2$ such that*

$$E(V_1, V_2) \geq 0.87856 \cdot \mathrm{OPT}.$$

Assuming a plausible complexity-theoretic conjecture (viz. the Unique Games Conjecture), the constant for this approximation algorithm for MAX-CUT is optimal.

Theorem 10. *There exists a polynomial time algorithm that colors a 3-colorable graph G using $\tilde{O}(n^{1/4})$ colors.*

The number of colors used in this algorithms is closed to optimal (for approaches based on SDP relaxations).